**Técnicas de Gravação e Mixagem de áudio**

**Apostila 2**

**Psicoacústica**

## O Ouvido Humano

O nosso processo auditivo está intimamente ligado às nossas funções de sobrevivência e tem padrões diferentes do de outras espécies de mamíferos. Podemos perceber sons com uma amplitude muito pequena (como o voo de um inseto, por exemplo), mas também percebemos outros sons com energia extrema (como o de um avião a jato). Também aprendemos, desde cedo, a interpretar os padrões sonoros que percebemos para aprimorar nosso senso de direção.

Nosso aparelho auditivo é complexo, e pode ser dividido nas seguintes partes:

**Orelha**: é o pavilhão auricular na parte externa do ouvido. As orelhas são duas, situadas nas partes laterais da nossa cabeça, e permitindo-nos uma orientação espacial. Os eventos sonoros, ao atingirem as nossas orelhas, que tem um formato similar ao de uma concha, sofrem uma série de reflexões e ressonâncias. Devido ao seu formato, as orelhas nos permitem captar sons vindos de diferentes direções, enviando-os ao conduto (ou canal) auditivo.

**Conduto Auditivo**: o tamanho e diâmetro do canal auditivo estão afinados para incrementar as frequências médias em até três vezes, desde que essas frequências estejam dentro dos limites de ressonância do canal. Essa característica nos permite perceber de forma mais intensa os sons que poderiam representar ameaças.

**Tímpano**: ao final do canal auditivo encontramos o tímpano, uma membrana que ocupa toda essa extremidade do canal, isolando-o da parte interna do nosso ouvido. Até chegar ao tímpano, o som ainda é percebido por movimentos do ar, mas, ao passar por essa membrana, o som é representado por vibrações na estrutura interna do ouvido. Após o

tímpano, temos o ouvido médio, que é uma região que contém o ar vindo das Trompas de Eustáquio, que se liga com a nossa boca. A necessidade de uma região com ar após o tímpano é prevenir que o próprio tímpano seja empurrado totalmente para dentro em razão de mudanças na pressão atmosférica. Assim, a Trompa de Eustáquio equaliza a pressão interna e externa do ouvido. Quando temos um resfriado, é possível que, com o acúmulo de secreção, haja um entupimento das Trompas de Eustáquio, dificultando a nossa capacidade natural de equalizar a pressão interna com a externa. Quando isso acontece, temos muita dificuldade de interpretar os sons. Devemos evitar viagens aéreas quando estivermos com essa condição, pois, durante um vôo, há uma pressurização forçada da cabine e podemos sofrer lesões.

**Ossículos**: o tímpano está conectado, dentro do ouvido médio, a três pequenos ossos móveis chamados de ossículos, que aumentam a proporção de movimento que se processou no tímpano, mais uma vez pela grandeza de três, antes de o som seguir para o ouvido interno. Acoplado aos ossículos, temos dois pequenos músculos que reagem à amplitude do som recebido. Quando esses músculos se contraem, eles diminuem a transmissão entre o tímpano e o ouvido interno. Esse é um movimento reflexo quase imediato a sons de intensidade extrema, contudo, algum tempo ocorre para ele reagir. Por essa razão, se estivermos muito próximos a um tiro de uma arma pesada, chegamos a ouvir um estrondo, mas, logo a seguir, temos um silêncio, que é resultado da ação desses músculos. Alguns filmes de ação se esquecem de que a reação a sons extremamente altos se processa dessa forma, e alguns dos seus efeitos sonoros se perdem ou são amenizados pela própria reação involuntária humana.

**Ouvido Interno/Cóclea**: o ouvido interno, ou a cóclea, tem o formato de um caracol e é preenchido por líquidos. Dentro da cóclea, temos uma série de membranas que atuam como se fossem tambores afinados para vibrar a uma determinada frequência, fazendo que esse órgão atue como

um analisador de frequências. Quando as vibrações atingem a cóclea, são convertidas em impulsos elétricos e transmitidas ao cérebro.

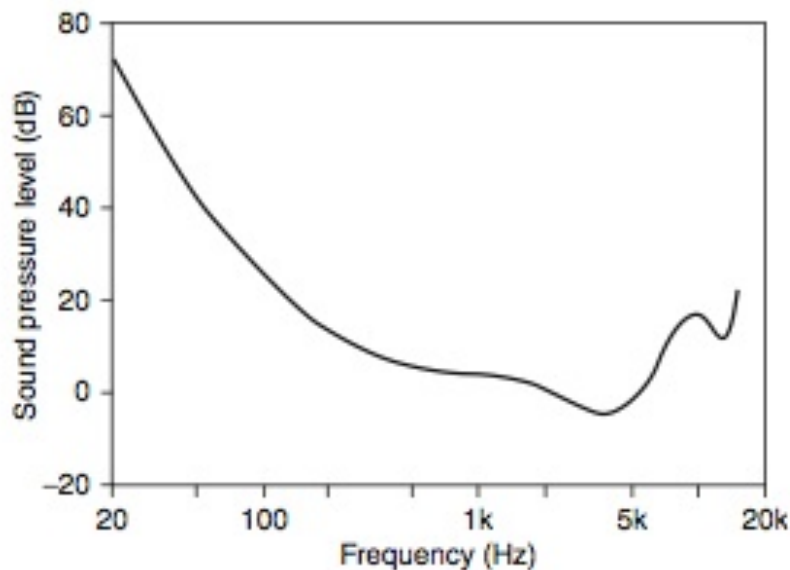
A audição deteriora naturalmente conforme envelhecemos. Esse fenômeno é registrado até em comunidades remotas, longe dos grandes centros urbanos. Hoje em dia, somos massacrados com uma infinidade de sons que não são naturais, com amplitudes que também não são facilmente encontradas na natureza, podendo acarretar a perda parcial ou até total da audição, se formos expostos por longos períodos a sons de intensidade exagerada. Por essa razão, encontramos muitas pessoas atualmente que tem problemas de perda de audição. Basicamente existem duas explicações para a perda de audição, quando ela ocorre: o envelhecimento e a exposição prolongada a sons com amplitudes extremas.

80db SPL é considerado o limite para a exposição a sons muito fortes. Abaixo dessa medida, ruídos não são considerados danosos ao aparelho auditivo. Já para sons medidos acima de 80db SPL, recomenda-se um limite que respeite a regra 3db de *volume x tempo*. Ou seja, 85db por no máximo 8 horas, 88db por no máximo 4 horas, etc.

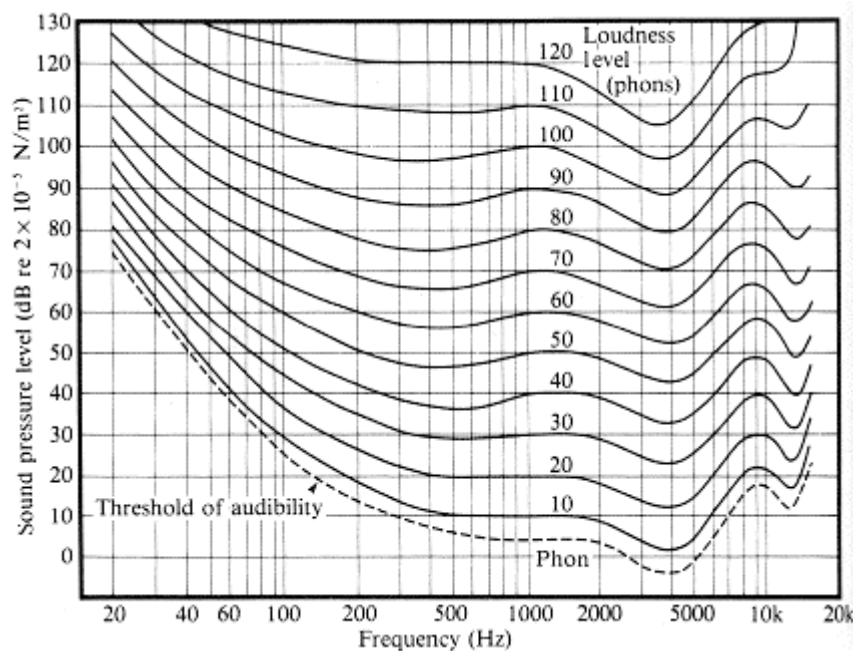
Na maioria das salas de cinema, os limites para volume são respeitados (será?). Porém, ambientes como *boîtes,* shows de rock, etc., são considerados muito mais perigosos que outros ambientes para a audição humana.

### *Loudness*

Nossa percepção auditiva não é homogênea para todas as frequências. Credita-se esse fato à evolução e à própria necessidade que o ser humano teria de se defender e de estar alerta aos perigos que o rondavam.

O gráfico abaixo mostra a percepção de frequências versus a energia mínima necessária para tornar o material audível.

Repare que o nosso pico máximo de percepção, onde necessitamos de menos amplitude para perceber as frequências, se situa entre 2-4 kHz. Esse padrão é considerado como uma média entre as pessoas, havendo é claro, indivíduos que têm sua audição fora desse padrão.

Podemos obter um gráfico que nos mostre as relações entre as várias frequências e os decibéis necessários para termos uma percepção de que estejam em um mesmo volume. A essa sensação damos o nome de intensidade subjetiva do som. Para sabermos qual é a diferença entre níveis de intensidade de frequências diferentes utilizamos uma unidade chamada fono. É importante não confundir o decibel com o fono, uma vez que o dB é usado para medir intensidades e potências e o fono é empregado nas medições de sonoridade. Para tanto, tomou-se como referencial a frequência de 1 kHz. Pesquisas foram feitas a partir da sensação de que outras frequências sejam

sentidas como tendo a mesma amplitude de 1 kHz a 20db SPL. A unidade que mede a percepção de *equal loudness* foi chamada de *phons*.

Através do gráfico, observe que na frequência de 1 kHz os valores em fonos coincidem com os de decibéis, mas o mesmo não ocorre em outras frequências. Repare que um som de 90 Hz que está 60 dB acima do nível zero, produz no ouvido a mesma intensidade subjetiva (40 fonos) que outro de 5 kHz 40 dB acima do nível zero. Um som de 300 Hz e 30 dB acima do nível zero, produz a mesma intensidade subjetiva (20 fonos) que outro de 3 kHz 14 dB acima do nível zero. Podemos notar que 100 Hz a 20 db SPL é um som inaudível e que, para começarmos a perceber 100 Hz necessitamos de algo próximo a 40 db SPL. Observamos também que não há, dentre todas as curvas formadas, nenhuma de resposta frequencial plana (*flat frequency response*), ou seja, nenhuma que trace uma reta.

*(Mostrar exemplo de áudio)*

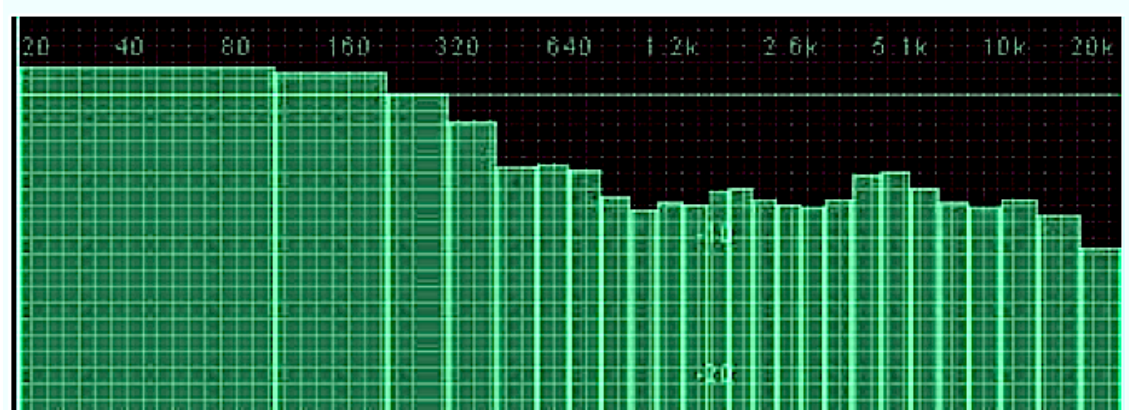
Todos esses fatores devem ser levados em conta quando tratamos um áudio para um filme, CD de música, ou até quando construímos salas de cinema, sabendo que devemos ter que valorizar as frequências baixas na reprodução do som. Antigos equipamentos de som continham uma chave que acionava um comando de nome *loudness*, que simplesmente mudava a equalização do áudio, dando um ganho maior nas baixas e altas frequências.

A escala de decibéis, logarítmica, reduz a nossa linearidade de modo que possamos trabalhar com a percepção - subjetiva - de uma forma mais objetiva. Experimentos feitos em percepção de amplitude revelaram que, para conseguirmos a sensação do dobro de volume, precisamos aumentar a nossa amplitude algo em torno de 6 a 10 db. Todavia, a cada 3 db estamos dobrando a energia aplicada a essa fonte sonora, mas para percebemos como um som com o dobro do volume, precisamos de aproximadamente 10 db.

Outro fator importante é o quanto de tempo é necessário para que tenhamos totalmente a sensação de *loudness,* uma vez que não percebemos mudança na amplitude imediatamente. Estudos indicam ser necessários pelo menos 30ms (milissegundos) para que possamos perceber alguma variação. Isso implica, em linguagem cinematográfica, em 8 frames de uma película a 24 fps (*frames* p/segundo).

*(Mostrar exemplo de áudio)*

**Espectro frequencial**

Podemos dividir, dentro dos limites de graves e agudos que compõem nossa capacidade perceptiva do espectro do audível, 24 regiões, ou bandas críticas (*critical bands*). O tamanho dessas bandas varia, sendo as mais largas nas regiões mais graves. Quando duas frequências com a mesma amplitude, que ocupam a mesma banda, soam simultaneamente, nossa percepção de *loudness* aumenta. Se as frequências, ainda com a mesma amplitude, mas em bandas diferentes, soarem juntas, a sensação de *loudness* será ainda maior.

Podemos concluir, então, que termos a sensação de *loudness* não depende apenas da frequência, mas do espectro (as duas frequências juntas) das frequências. De uma maneira geral, podemos dizer que quanto maior o espectro frequencial de um som, maior será a sensação de *loudness*. A técnica de adicionar frequências, na pós-produção de um áudio, para enfatizar um som específico, é muito utilizada tanto no cinema, como na produção musical.

**Mascaramento frequencial**

Os sons com volume mais baixo tendem a ser encobertos por sons com maior volume, especialmente se as frequências dos sons estiverem próximas, criando o que chamamos de *mascaramento frequencial* (*frequency masking*). De modo a evitar que isso ocorra inadvertidamente, recomenda-se espalhar as frequências importantes pelo espectro sonoro audível. Compositores de trilhas e sonoplastas têm de ter isso em mente quando estão escrevendo a música ou sonorizando um filme. Uma sugestão para o uso do efeito de *frequency masking* ocorre quando temos uma cena em que o som ambiente não está satisfatório. Pode-se nesse caso, colocar outra pista de ruídos com outro som ambiente que seja mais apropriado e manter essa segunda opção com maior amplitude, *mascarando* o som ambiente original.

*(mostrar exemplo de áudio)*

**Dissimulação temporal (*temporal masking*)**

Um som com maior amplitude pode mascarar outro mesmo que estes não ocorram no mesmo momento. Um exemplo recorrente no cinema seria o que acontece com o som de um tiro, que esconde todos os outros sons assim que é percebido, sendo que a sensação que temos é a de que o tiro mascara os outros sons por algum tempo após o tiro ser disparado. Curiosamente, esse efeito de dissimulação ocorre também de forma retroativa, pois um tiro pode encobrir a nossa percepção de outro som até 10ms antes de se iniciar. Na verdade, é a nossa percepção que cria esse *atraso*, que seria o tempo que leva para nosso cérebro processar toda a informação. De qualquer maneira, pode-se utilizar esse recurso em pós-produções de áudio quando temos alguma

falha ou silêncio indesejado. Basta colocar um som forte, podendo ser da própria trilha musical (ataque forte de orquestra, guitarra, bateria, etc.), para criar uma dissimulação temporal, ocultando a falha.

O fato de que fenômenos (*frequency* e *temporal* masking) dessa natureza acontecem está por trás dos algoritmos de compressão de áudio digital, pois como tais sons não serão percebidos pelo ouvinte, não faz sentido arquivá-los digitalmente.

*(mostrar exemplo de áudio)*

***Pitch Shifting***

Vimos anteriormente que os sons harmônicos (com *pitch*) são caracterizados pela identidade da amplitude dos seus harmônicos. A voz humana, mesmo quando não está associada ao canto, possui um timbre definido – e, consequentemente, uma fundamental e seus harmônicos. Podemos equalizar a voz enfatizando graves, médios ou agudos, mas também podemos utilizar a técnica de mover a fundamental, mudando o *pitch* do que é falado e gerando outras características para o sinal, como, por exemplo, dotar uma voz masculina de características femininas (vozes femininas tem um *pitch* mais agudo). Também é possível dar a uma voz características não naturais, se movermos seu *pitch* para uma região grave demais. A isso damos o nome de *Pitch Shifting*.

**Percepção Espacial**

O ser humano percebe o som, rotineiramente, em três dimensões, possibilitando que tenhamos uma percepção espacial que se completa com o uso da visão. Um dos desafios do áudio em cinema é como representar essas dimensões, com o material sonoro, sem comprometer a lógica visual, já que, ao contrário da nossa visão no cinema (em que a câmera tem autonomia do que nos mostrar) o som é invasivo. Por exemplo, o som direto pode captar sons que não estejam em quadro, confundindo a percepção da platéia. Alguns elementos nos ajudam a entender esse fenômeno:

Sons transientes: são sons de curta duração, mas com presença (ex. estalar de dedos, batidas fortes em uma porta, etc.). Os sons transientes são excelentes ferramentas para que tenhamos a noção da localização de um som.

*(mostrar exemplo de áudio)*

Efeito Precedente: é a atenção que damos ao som que chega primeiro aos nossos ouvidos. Essa sensação só é apagada se outro som chegar com amplitude maior logo a seguir.

Visão: tem que estar em conjunção com a audição. Embora a visão seja predominante na questão da localização espacial, é necessário que não existam discrepâncias entre o visual e o auditivo, pois isso pode ser causa de dissonâncias cognitivas.

Localização em 3 dimensões: percebemos melhor o som na sua horizontalidade provavelmente porque nossos ouvidos estão dispostos dessa forma, nos dois lados da nossa cabeça. Por exemplo, um som vindo do lado direito chegará ao nosso ouvido direito antes de chegar ao esquerdo, aonde chega por difração. Dizemos então que o ouvido esquerdo é a sombra acústica (*acoustic shadow*) da nossa cabeça. Frequências graves - ondas maiores - chegam com facilidade ao segundo ouvido, reduzindo o efeito de *sombra*. Já com as frequências agudas, para as quais a cabeça é um objeto relativamente grande devido ao tamanho das ondas, que são pequenas, o nível de volume que chega ao segundo ouvido é substancialmente menor.

Para a percepção de profundidade, temos que trabalhar com elementos menos distintos, como o volume e o brilho da fonte sonora (a curtas distâncias percebemos mais claramente as frequências agudas), o efeito Doppler para objetos em movimento, e os padrões de reverberação de um ambiente (os mais amplos tendem a ter uma taxa de reverberação maior). Podemos trabalhar essa percepção na narrativa cinematográfica, de forma a, por exemplo, conferir à voz do narrador um nível de reverberação excessivo (que nos leva a crer que se trata de um pensamento do personagem), ou então, pelo contrário, retirar da voz do narrador toda a reverberação similar à da cena apresentada, sugerindo assim a percepção de que ele está em *off* ou fora da cena.

Outra possibilidade, muito utilizada em filmes de época, ocorre quando usamos, por exemplo, uma música existente na época, mas, ao invés de usarmos uma gravação recente, pegamos um equipamento de reprodução da própria época em que se passa o filme e gravamos a reprodução do som gerado por esse equipamento antigo, não apenas localizando a música no tempo, mas também toda a "atmosfera" sonora dessa época.

Muitas vezes, mesmo estando em um ambiente sonoro com muita reverberação em que vários grupos de pessoas falam simultaneamente, somos capazes de entender os outros e também de nos fazer por eles entender. Isso acontece devido muitos fatores, como a ajuda visual auxiliada por alguma leitura labial e o conhecimento prévio das pessoas sobre o tópico da conversa, o que ajuda a completar um diálogo incompleto, etc. Quando temos de gravar diálogos em situações similares, temos desafios pela frente. Muitas vezes podemos utilizar microfones de lapela nos principais personagens, mas outras vezes, isso pode não ser possível. Existem algumas opções para essa situação, como a filmagem da cena apenas com os personagens principais falando, enquanto os figurantes fazem mímica labial. Também se recomenda a gravação de *wild tracks* de som ambiente (sem diálogos) e de takes de áudio com os personagens principais falando o diálogo em *off*, sem sincronismo com a imagem, para termos opções, se o áudio original tiver alguns defeitos.

**Objeto Sonoro**

Entendemos objeto sonoro como um som que pode ser identificado entre outros sons. Por exemplo, podemos ter um diálogo com dois atores. Nesse caso, cada ator é um objeto sonoro. No som de um filme, temos vários elementos, como os sons ambientes, os efeitos, os diálogos, a música, etc. Devemos agrupar os sons como objetos sonoros sempre que tiverem características semelhantes, e trabalhar o áudio para que tais objetos sejam facilmente identificáveis. O mesmo vale para a gravação de instrumentos musicais. Cada instrumento (ou voz) deve ser entendido como um objeto sonoro distinto.

Podemos trabalhar o áudio de forma a facilitar o entendimento do objeto sonoro atentando para os seguintes elementos:

Timbre: sabemos que o timbre é a identidade da amplitude dos harmônicos, portanto, se a voz de um ator ou de um instrumento da trilha não soa com as características de deveriam, podemos mexer na equalização para definirmos o objeto.

Sons transientes: um som com essa característica deve soar com clareza (um tiro, por exemplo), caso contrário pode perder a sua função.

Ênfase na frequência fundamental: se um objeto sonoro importante, como o tiro do exemplo anterior, não está com a clareza necessária, podemos dobrar a fundamental do som em outra pista, enfatizando o objeto sonoro.

Localização do objeto sonoro: se vários sons deveriam estar no mesmo ambiente, mas estão claramente em ambiências diferentes, podemos tentar, controlando a reverberação na pós-produção, localizá-los via ambiência.

Contrastes: podemos utilizar objetos sonoros diferentes para a transposição de cenas (de uma tomada interna para uma externa, por exemplo). Podemos, ao final de uma cena interna, colocar o som ambiente da cena externa seguinte. A platéia irá estranhar, mas essa quebra já coloca o espectador no próximo ambiente. Pode-se utilizar um *crossfade* entre as duas cenas, minimizando o contraste.

Ritmo: elementos da música, dos efeitos ou da sonoplastia podem criar a sensação de ritmo, com pulsações claramente perceptíveis. O ritmo, quando claramente estabelecido, é um objeto sonoro, e deve se entendido como tal. Esse é o caso de passos, palmas, pingos de torneira, etc. Podemos acelerar e diminuir o ritmo, criando sensações na platéia.

Similaridade: sons com o mesmo *pitch, loudness,* timbre e localização, podem formar um só objeto sonoro.

Alterações comuns a dois ou mais sons: se fizermos alterações de volume, timbre, ambiência, etc. em um ou mais sons que originalmente seriam percebidos como sons distintos, estes passam a ser considerados como um único objeto sonoro. Um orquestrador, por exemplo, irá misturar dois ou

mais instrumentos para criar outra identidade sonora, gerando apenas um objeto sonoro que, do ponto de vista de sua gravação, edição e mixagem, corresponde à soma dos instrumentos.

Cortes musicais: a música, em muitos casos, pode ser (e é) um objeto sonoro que, para sofrer cortes, deve seguir uma coerência especificamente musical. Essa continuidade musical, no caso de uma trilha para filme ou peça de teatro, pode não servir ao propósito da cena. Nesse caso, temos de introduzir outro elemento, como um efeito sonoro de grande amplitude, por exemplo, para que a platéia não sinta o problema do corte musical.

Quantidade de sons: segundo Walter Murch, o *sound designer* do filme *Apocalipse Now*, de Francis F. Coppola, conseguimos prestar atenção a, no máximo, três sons simultaneamente. Portanto, quando estivermos sonorizando, não devemos colocar elementos sonoros demais, pois eles correm o risco de nem serem percebidos. O mesmo vale para um arranjador e/ou compositor, se ele colocar elementos demais em sua peça musical, esta terá seu entendimento comprometido.

Visual *Syncronization*: tudo que é visto na tela, queremos ouvir nos alto-falantes. Ou seja, se um carro passa na tela, temos a expectativa de que um ruído de carro esteja no som do filme. O mesmo acontece ao contrário, pois, se ouvimos o ruído de um carro, esperamos a imagem do automóvel correspondente na tela.